



**Sujet de stage: Dynamique d'expression du génome en lien avec la diversité architecturale de la panicule chez 3 espèces de riz.**  
*méthodes de quantification de transcrits sans alignement*

Mots clés : riz, panicule, transcriptomique, kmers, mapping, quantification, cinétique, bioanalyse

Les travaux de l'équipe EDI, de l'UMR DIADE (Diversity - Adaptation - plant DEvelopment), sont centrés sur la biologie de la reproduction du riz, appartenant au genre *Oryza*. Ce dernier comprend deux espèces cultivées, *O. sativa* et *O. glaberrima*, domestiquées de façon indépendante à partir de leurs ancêtres sauvages (*O. rufipogon* et *O. barthii*). Chez le riz, la complexité de branchement de l'inflorescence (ou panicule) est un caractère directement lié au rendement en grain et on observe une grande variabilité inter-et intra-spécifique<sup>1</sup>.

Dans le but d'identifier des facteurs moléculaires liés à la diversité d'architecture de la panicule, une étude comparée de la dynamique d'expression du génome a été initiée sur 9 stades de différenciation de la panicule et chez 3 espèces de riz à architecture paniculaire. Ce travail permettra de suivre la trajectoire moléculaire (dynamique) de développement de la panicule et d'associer des différences de dynamique d'expression à la variabilité architecturale de la panicule observée chez ces espèces.

Certaines limites méthodologiques ont été détectées, concernant la quantification parallèle des transcrits chez plusieurs espèces, qui est conditionnée par la qualité du mapping (choix de l'outil et du génome de référence utilisé). Même si disponible pour le riz, le recours à un génome ou à un transcriptome de référence est une source de biais dans l'analyse conventionnelle des données RNA-seq. Ces méthodes ne tiennent pas compte des nombreux ARN produits dans différentes conditions, et il y a encore de nombreuses espèces pour lesquelles aucun génome ou transcriptome de référence n'est disponible.

L'objectif du stage est d'explorer des alternatives aux méthodes classiques de quantification des transcrits en exploitant une approche sans alignement par des kmers, des sous-séquences de taille  $k$  extraits directement des reads. Pour cela, plusieurs outils existent et pourront être testés (KE-dupl<sup>2</sup>, GECKO<sup>3</sup>, iMOKA<sup>4</sup>). Dans le cadre de ce stage, une quantification des transcrits via une méthode sans alignement sera réalisée pour aboutir à la comparaison des profils des cinétiques entre les 3 espèces. Les profils et modules de co-expression de gène au cours du développement seront comparés entre les 3 espèces. Ces résultats pourront être comparés aux résultats obtenus par des méthodes classiques.

Compétences requises :

Maîtrise des systèmes Unix

Programmation en R et Python  
Bases en statistique  
Curiosité scientifique

Stage encadré par :

Hélène Adam [helene.adam@ird.fr](mailto:helene.adam@ird.fr) , Chargé de recherche  
James Tregear [james.tregear@ird.fr](mailto:james.tregear@ird.fr) , Directeur de recherche  
Julie Orjuela ([julie.orjuela@ird.fr](mailto:julie.orjuela@ird.fr) ), Bioinformaticienne

UMR DIADE - <http://diade.ird.fr/>  
IRD - Institut de Recherche pour le Développement  
Montpellier

Bibliographie :

1. Harrop TWR., Mantegazza O., Luong AM., Béthune K., Lorieux M., Jouannic S., Adam H. (2019) A set of AP2-like genes is associated with inflorescence branching and architecture in domesticated rice. *Journal of Experimental Botany* 70: 5617–562.
2. Audoux, J. *et al.* DE-kupl: exhaustive capture of biological variation in RNA-seq data through k-mer decomposition. *Genome Biol.* **18**, 243 (2017).
3. Thomas, A. *et al.* GECKO is a genetic algorithm to classify and explore high throughput sequencing data. *Commun. Biol.* **2**, 1–8 (2019).
4. Lorenzi, C. *et al.* iMOKA: k-mer based software to analyze large collections of sequencing data. *Genome Biol.* **21**, 261 (2020).